

[BoB_7th_DF]김성수_Parser

Gmail Parsing, URL Authenticate, Image Download, Draw Gmap

/* 개요 */

개발환경 및 설정

요구사항 및 결과

구현방법

/* 첨부파일 */

결과파일

소스코드

실행파일

개발 환경 및 설정

1. 개발환경

OS : Windows 7 Ultimate

interpreter : python 2.7

2. 핵심 Module 2개 설치 및 설정

가. gmail

<https://github.com/charlierguo/gmail>

pip support not yet implemented

python 2.7 is needed

GMail for Python

A Pythonic interface to Google's GMail, with all the tools you'll need. Search, read and send multipart emails, archive, mark read/unread, delete emails, and manage labels.

This library is still under development, so please forgive some of the rough edges

Heavily inspired by Kriss "nu7hatch" Kowalik's GMail for Ruby library

Author

- [Charlie Guo](#)

Installation

For now, installation is manual (pip support not yet implemented) and the only requirement is to use Python 2 (2.7+ to precise):

```
git clone git://github.com/charlierguo/gmail.git
```

나. gmpplot

pip2 install gmpplot

라이브러리 __init__ 함수의 self.coloricon 경로 수정 필요(drop pin 관련 부분)

```
gmpplot.py
class GoogleMapPlotter(object):
    def __init__(self, center_lat, center_lng, zoom, apikey=''):
        self.center = (float(center_lat), float(center_lng))
        self.zoom = int(zoom)
        self.apikey = str(apikey)
        self.grids = None
        self.paths = []
        self.shapes = []
        self.points = []
        self.heatmap_points = []
        self.radpoints = []
        self.gridsetting = None
        # self.coloricon = os.path.join(os.path.dirname(__file__), 'markers/%s.png')
        self.coloricon = "http://www.googlemapsmarkers.com/v1/%s/"
        self.color_dict = mpl_color_map
        self.html_color_codes = html_color_codes

    @classmethod
    def from_geocode(cls, location_string, zoom=13):
        lat, lng = cls.geocode(location_string)
        return cls(lat, lng, zoom)
```

요구사항 및 결과

1. 요구사항

- 랜덤한 시간에 이메일로 특정 이미지파일의 단축URL이 전송됩니다
- 이메일에 자동으로 접속하고 다운로드받은 이메일 첨부파일을 정리하는 프로그램을 만드시오 (파이썬 언어로 작성할 것)

Q1) 프로그램은 한번 실행되면 "fl0ckfl0ck@hotmail.com"로부터 수신된 이메일의 본문에서 단축URL을 파싱하여 업로드된 이미지파일을 다운로드하는 기능을 구현할 것(20점) (○)

Q2) py2exe를 이용하여 윈도우 실행파일로 만들 것(20점) (○)

Q3) 윈도우 스케줄러를 이용하여 매일 11:50, 23:50 실행파일이 동작하도록 만들 것(10점) (○)

Q4) 프로그램이 실행되면 "YYYY-MM-DD"의 형태로 디렉토리를 만들고 그날의 결과물을 모두 저장할 것(결과물 : 각 이미지파일, csv 파일, 20점) (○)

Q5) 이미지파일의 EXIF 정보를 파싱하여 GPS 데이터를 추출한 뒤 구글맵으로 표현할 것 (10점) (○)

Q6) 위 과제 모두 해결 후 selenium 라이브러리를 이용하여 위 과제를 해결 할 것(20점) (X)

2. py2exe를 이용하여 윈도우 실행파일로 만들 것

가. 윈도우 64비트 환경에서는 py2exe 가 onefile 로 만드는 기능을 제공하지 못한다.

SOURCEFORGE					
py2exe-0.6.9.win32-py2.5.exe	2008-11-16	184.6 kB	170		i
py2exe-0.6.9.win32-py2.4.exe	2008-11-16	184.1 kB	4		i
py2exe-0.6.9.win32-py2.3.exe	2008-11-16	176.7 kB	8		i
py2exe-0.6.9.win32-py2.6.exe	2008-11-16	201.2 kB	28		i
py2exe-0.6.9.win32-py2.7.exe	2008-11-16	201.2 kB	1,564		i
py2exe-0.6.9.win64-py2.5.amd64.msi	2008-11-16	172.0 kB	30		i
py2exe-0.6.9.win64-py2.6.amd64.exe	2008-11-16	220.1 kB	18		i
py2exe-0.6.9.win64-py2.7.amd64.exe	2008-11-16	220.1 kB	436		i
py2exe-0.6.9.zip	2008-11-16	149.7 kB			
Totals: 9 Items		1.7 MB			

Get latest updates about Open Source Projects, Conferences and News.

Sign Up

No, Thank you

나. pyinstaller 를 통해 윈도우 실행파일로 구현

pyinstaller -F --onefile [codeName].py

```
관리자: C:\Windows\System32\cmd.exe
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

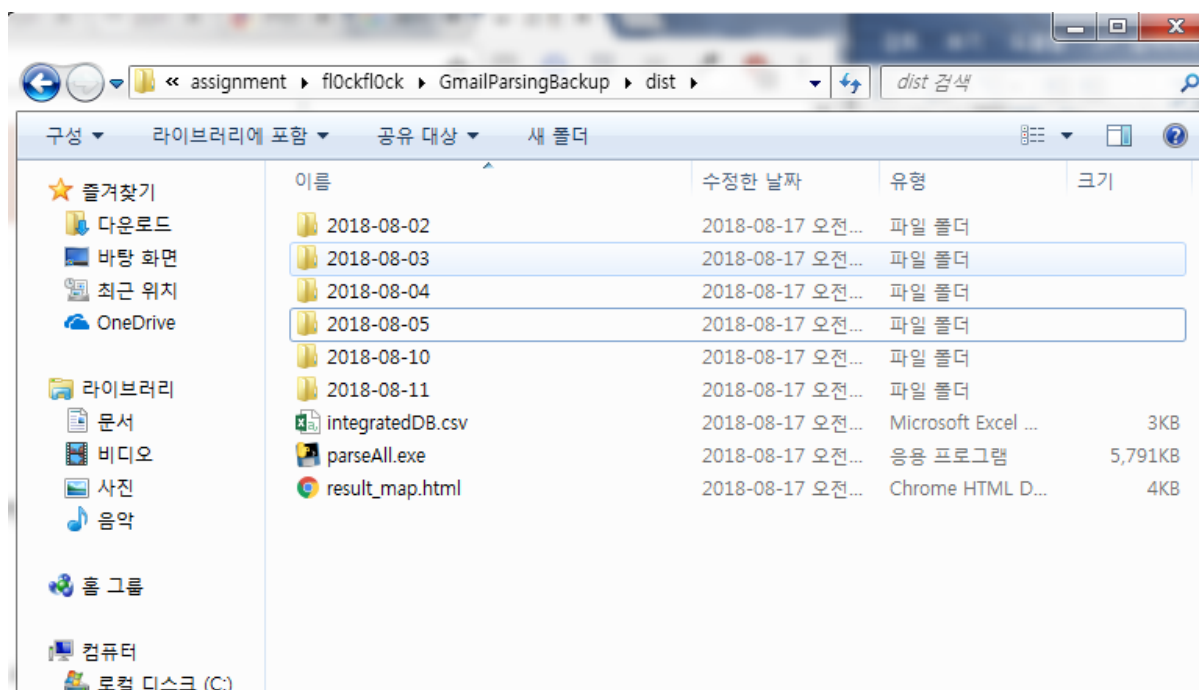
D:\wobob_forensic_7th\assignment\wf10ckf10ck\GmailParsingTest>pyinstaller -F --onefile GmailParser.py
201 INFO: PyInstaller: 3.3.1
203 INFO: Python: 2.7.15
203 INFO: Platform: Windows-7-6.1.7601-SP1
203 INFO: wrote D:\wobob_forensic_7th\assignment\wf10ckf10ck\GmailParsingTest\GmailParser.spec
210 INFO: UPX is not available.
212 INFO: Extending PYTHONPATH with paths
['D:\wobob_forensic_7th\assignment\wf10ckf10ck\GmailParsingTest',
'D:\wobob_forensic_7th\assignment\wf10ckf10ck\GmailParsingTest']
213 INFO: checking Analysis
12525 INFO: Building PEX archive out00-PEX.pkg
12547 INFO: Redirecting Microsoft.UC90.CRT version (9, 0, 21022, 8) -> (9, 0, 30729, 6161)
12565 INFO: Building PKG (CArchive) out00-PKG.pkg completed successfully.
12572 INFO: Bootloader c:\python27\lib\site-packages\PyInstaller\bootloader\Windows-64bit\run.exe
12572 INFO: checking EXE
12574 INFO: Building EXE because out00-EXE.toc is non existent
12576 INFO: Building EXE from out00-EXE.toc
12578 INFO: Appending archive to EXE D:\wobob_forensic_7th\assignment\wf10ckf10ck\GmailParsingTest\dist\GmailParser.exe
12589 INFO: Building EXE from out00-EXE.toc completed successfully.

D:\wobob_forensic_7th\assignment\wf10ckf10ck\GmailParsingTest>
```

다. 2가지 버전의 .exe 파일

1) parseAll.exe

fl0ckf10ck 로부터 온 이메일을 전부 읽어서 처리. 독립적으로 실행가능

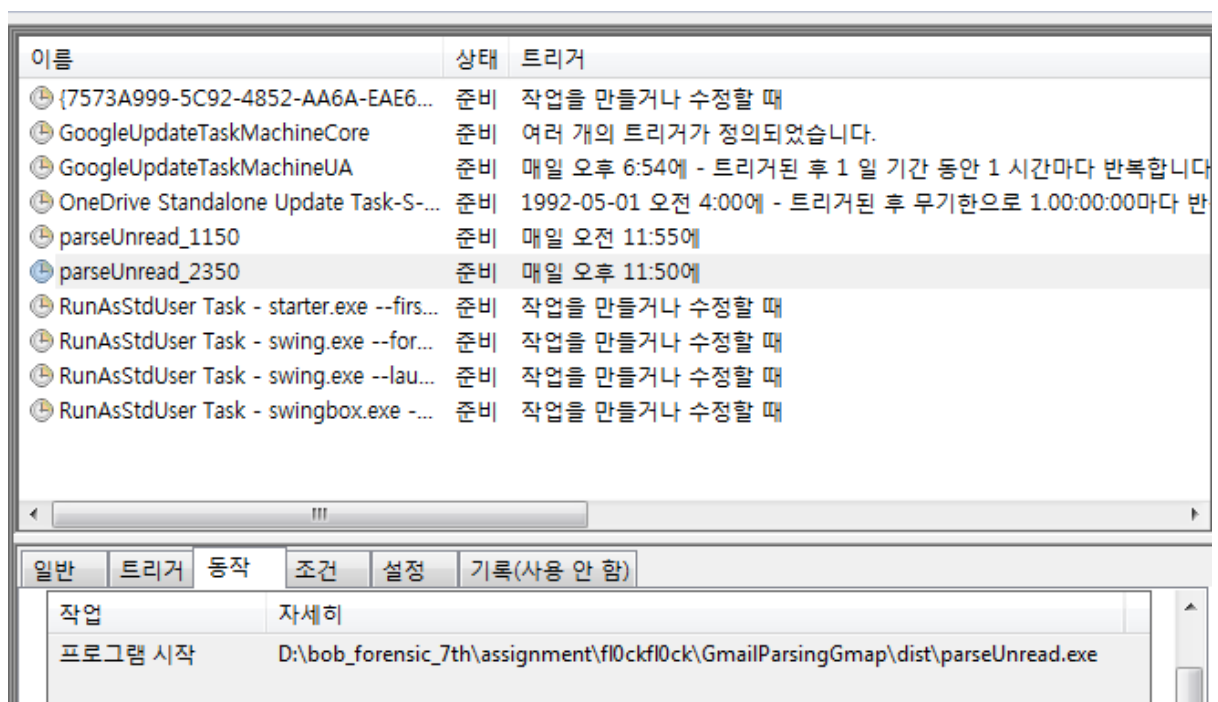


2) parseUnread.exe

f10ckf10ck 로부터 온 이메일 중, 읽지 않은 것을 읽어서 처리.

윈도우 작업스케줄러에 의해 parseUnread 가 꾸준히 실행되어야 한다. integratedDB 를 생성하는데 있어, 이전의 처리데이터를 사용하기 때문에, 한번 읽은 이메일의 처리데이터가 같은 경로 내에 있어야 한다. integratedDB.csv 는 parseUnread.exe가 실행될 때마다 갱신되며, result_map.html 을 만드는데 반영이 된다.

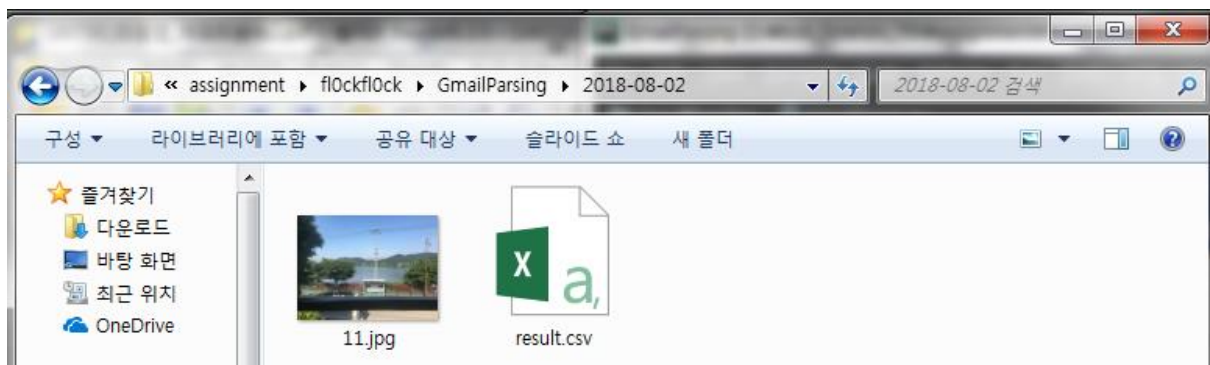
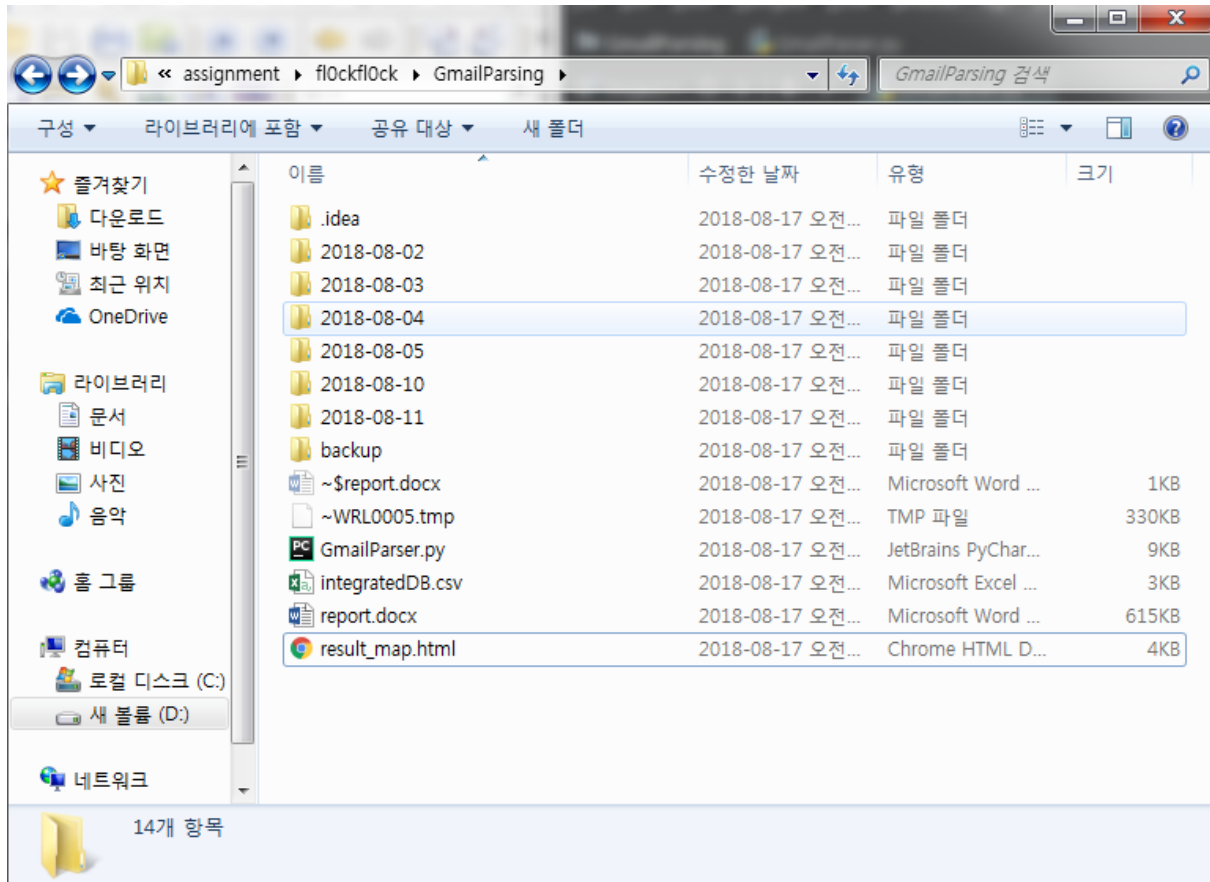
3. 윈도우 스케줄러를 이용하여 매일 11:50, 23:50 실행파일이 동작하도록 만들 것

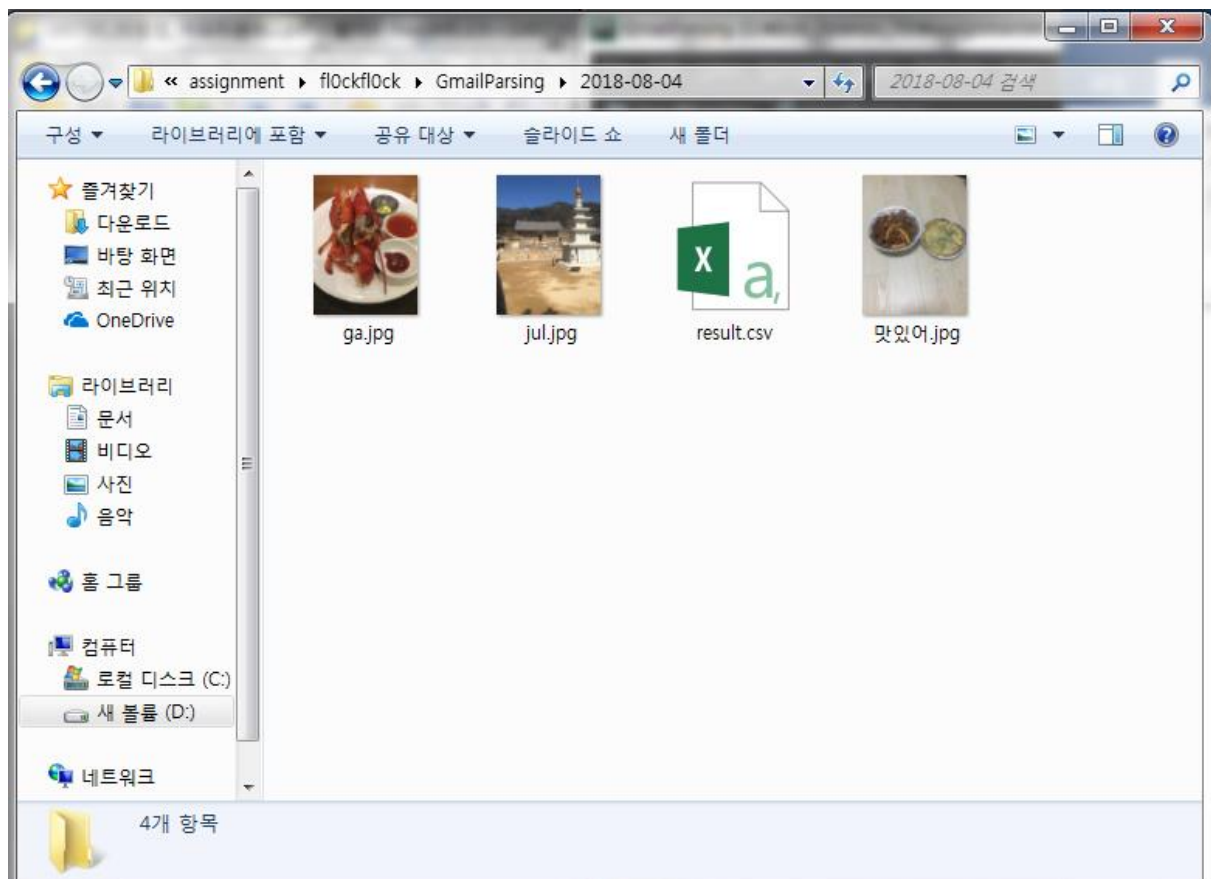
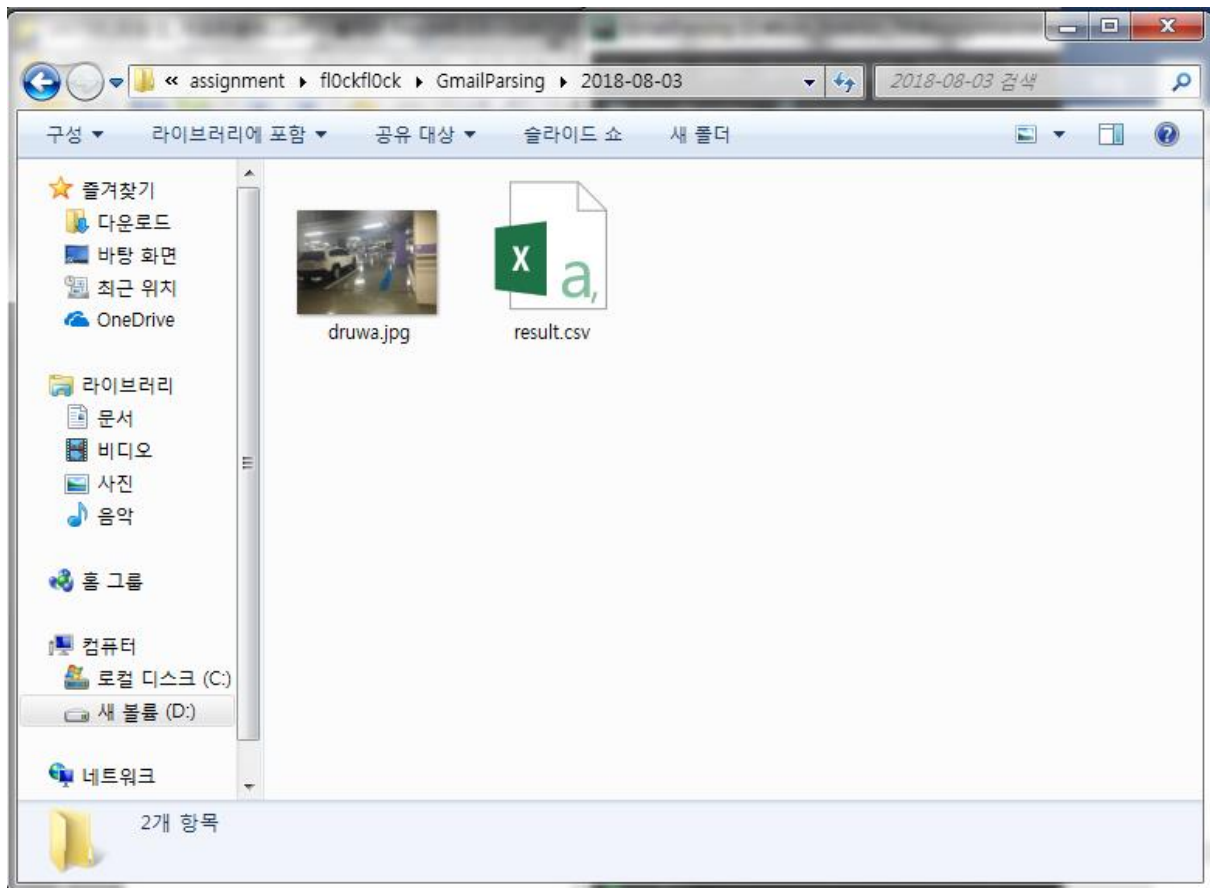


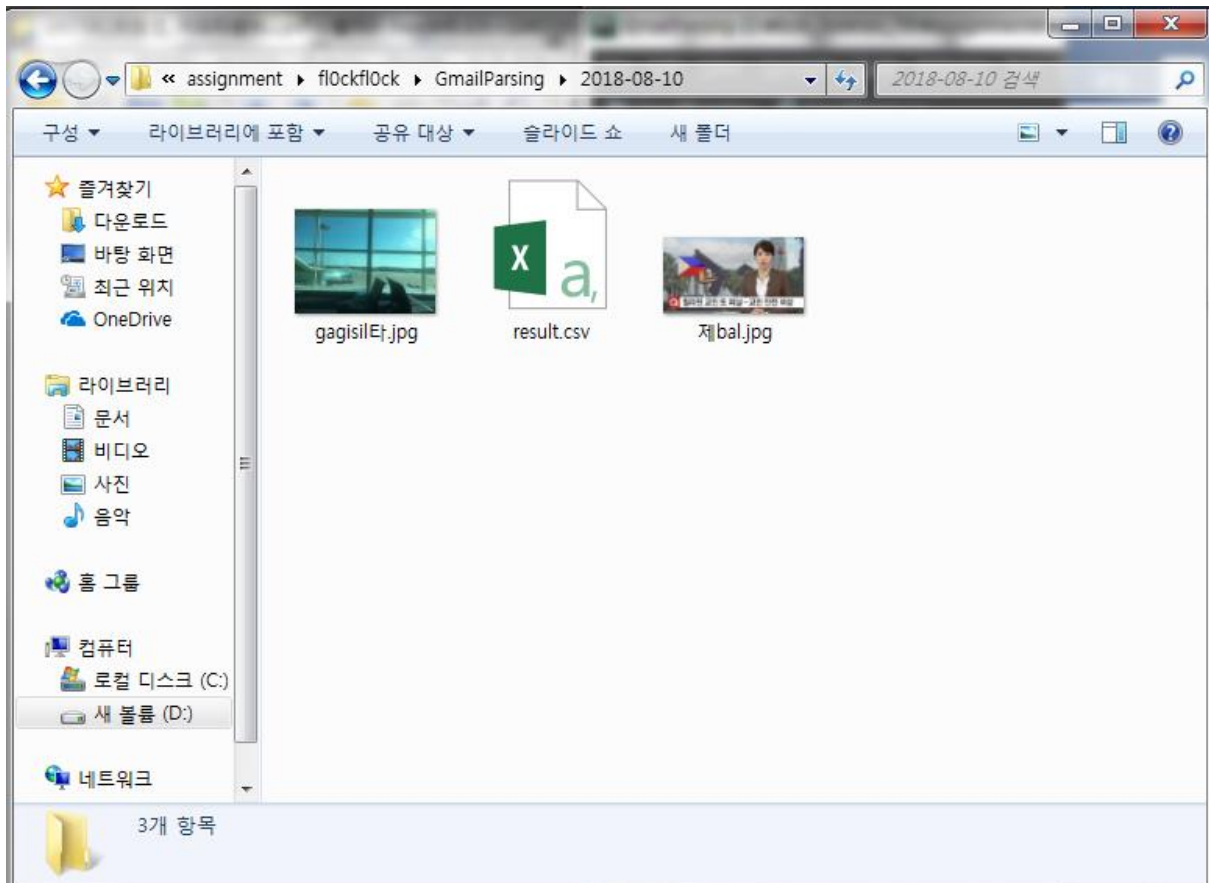
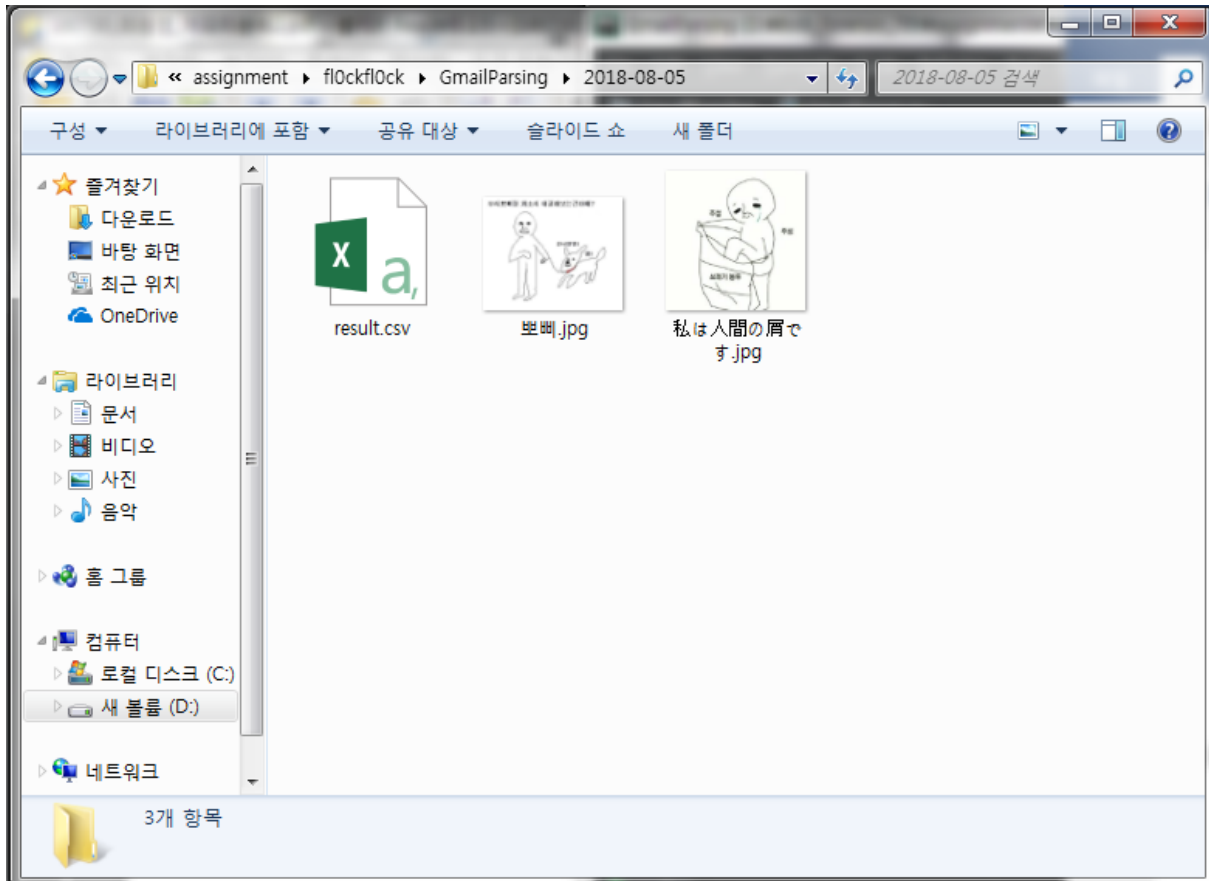
4. 실행 후 생성 파일 관련

4-1) 프로그램은 한번 실행되면 “fl0ckfl0ck@hotmail.com”로부터 수신된 이메일의 본문에서 단축URL을 파싱하여 업로드된 이미지파일을 다운로드하는 기능을 구현할 것

4-2) 프로그램이 실행되면 “YYYY-MM-DD”의 형태로 디렉토리를 만들고 그날의 결과물을 모두 저장할 것

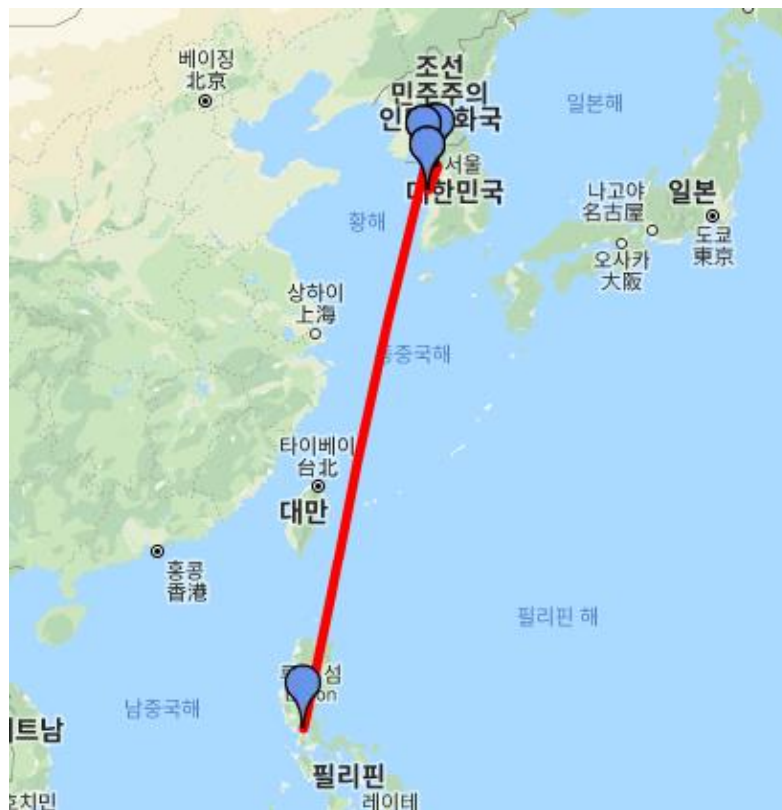
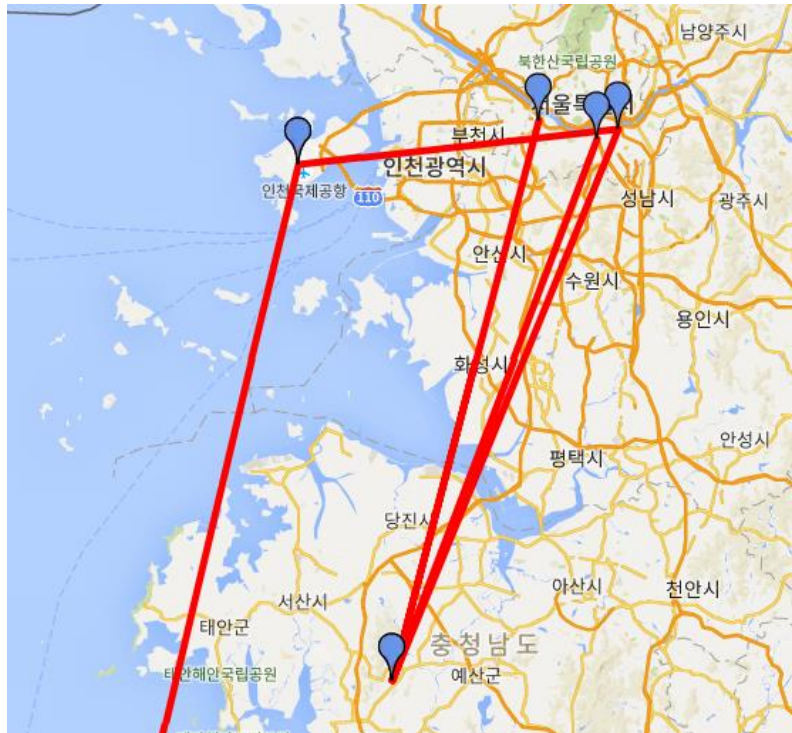






	integratedDB.csv	
1	2018-08-02 11:23:36	http://bit.ly.kr/3uA8 http://flockflock.info/11.jpg 11.jpg 0 0 4d7ab511771b7a532de3cd83de2ada8 f5863bae5370821782879f76f1d378ae8f
2	2018-08-02 11:23:36	http://flockflock.info/11.jpg 11.jpg 0 0 4d7ab511771b7a532de3cd83de2ada8 f5863bae5370821782879f76f1d378ae8f
3	2018-08-03 14:24:01	https://bit.ly/2QWwQR http://flockflock.info/dm/1.jpg 1.jpg 37.98489 127.018715 8a9f31422ca6545c6f8091807142d02 08fd0dc38e3e
4	2018-08-04 06:13:46	https://bit.ly/2MfPMhY http://flockflock.info/jul.jpg jul.jpg 36.662941 126.622589 8d9626538963f49bfa4c83566da0 19e09a58d6672797dc2
5	2018-08-04 10:47:36	http://bit.ly/2MfPMhY http://flockflock.info/ga.jpg ga.jpg 37.527790 126.904610 8bc2f799e3569e35679dc5c3936b5 276eb281bf4ae0f282ce2
6	2018-08-04 10:47:36	https://bit.ly/2MfPMhY http://flockflock.info/jul.jpg jul.jpg 36.662941 126.622589 8d9626538963f49bfa4c83566da0 19e09a58d6672797dc2
7	2018-08-05 11:24:24	http://bit.ly/2NkZE http://flockflock.info/gu/gu.jpg gu.jpg 37.53062 127.058647 4641e3242ce0efb0db62d323ab326 5365b88da062
8	2018-08-05 13:50:08	https://hoy.kr/SvCC http://flockflock.info/산/산.jpg 산.jpg 0 c99c871444c619a9cf9c5e4b39f748 f8d811667f7f80ac53423b1c79de9c913263c
9	2018-08-05 13:50:58	https://hoy.kr/MSDB http://flockflock.info/私/私の顔です.jpg 私の顔です.jpg 0 0 0a794cc0b49e2b767e87cf083256027 744928f2c2126566d0
10	2018-08-10 07:53:54	https://hoy.kr/Vc61 http://flockflock.info/ga/ga.jpg ga.jpg 37.468190 126.448697 0eb8b483b7a01de504cf382a 255c1312027d62388e
11	2018-08-10 07:54:37	https://hoy.kr/Vc61 http://flockflock.info/ga/ga.jpg ga.jpg 37.468190 126.448697 0eb8b483b7a01de504cf382a 255c1312027d62388e
12	2018-08-11 02:26:19	https://hoy.kr/SDB3 http://flockflock.info/dirty and rainy.jpg dirty and rainy.jpg 14.578697 128.976156 c6281b85067b0c8b2ba0885f40b124fe

5. 이미지파일의 EXIF 정보를 파싱하여 GPS 데이터를 추출한 뒤 구글맵으로 표현할 것



구현방법

1. gmail 모듈을 통하여, f10ckf10ck 로부터 온 이메일 중 읽지 않은 것의 데이터를 받아온다.
2. 이메일의 날짜를 파싱하여, YYYY-MM-DD 디렉토리를 만든다.
3. 이메일의 본문의 내용 중, 정규표현식을 통하여 <https://~.~/~> 형태의 문자열을 추출하여, 한글 자씩 대입하며 HTTP 응답코드가 200인 것을 찾는다.
4. 응답코드가 200인 것의 원래 URL 에 f10ckf10ck.info 가 포함되어 있는지 확인한다.
5. f10ckf10ck.info 가 포함되어 있으면, YYYY-MM-DD 폴더에 이미지를 다운로드 한다.
6. 이와 동시에, YYYY-MM-DD 폴더에 result.csv 를 만들고 이미지와 관련된 정보를 추가한다.
7. 위 과정을 읽지않은 이메일의 내용마다 반복한다.
8. 위 과정이 끝나면, 각각의 YYYY-MM-DD 폴더의 result.csv 를 종합하여 integratedDB를 만든다.
9. integratedDB를 읽어서 GPS 관련 정보를 추출하고, gmplot 모듈을 통하여 지도상에 표기한다.